



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2015

Distributional typology: statistical inquiries into the dynamics of linguistic diversity

Bickel, Balthasar

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-109110>

Book Section

Accepted Version

Originally published at:

Bickel, Balthasar (2015). Distributional typology: statistical inquiries into the dynamics of linguistic diversity. In: Heine, Bernd; Narrog, Heiko. The Oxford handbook of linguistic analysis, 2nd edition. Oxford: Oxford University Press, 901 - 923.

Distributional Typology: statistical inquiries into the dynamics of linguistic diversity^{*}

Balthasar Bickel
University of Zürich

Abstract: Over the past two decades, linguistic typology has been moving increasingly away from its original goal of classifying languages into ideal types that would be constrained by categorical universals. What has been emerging as a new paradigm instead starts from the distribution of structures in the world, asking “what’s where why?” I present here a concrete approach to this question, called ‘Distributional Typology’. The approach starts from causal theories on the forces that affect language change, from processing preferences to the historical contingencies of language contact. The predictions of these theories can then be tested against fine-grained matrices of cross-linguistic diversity, using statistical methods for estimating diachronic trends from synchronic distributions.

Over the past two decades, linguistic typology has been moving increasingly away from its original goal of classifying languages into ideal types that would be constrained by categorical universals. What has been emerging as a new paradigm instead starts from asking “what’s where why?”: *What* linguistic structures are there in human languages, and how can we compare them? *Where* do we find these structures, i.e. are they areally or genealogically restricted, or are they universally preferred or dispreferred? *Why* do we find the structures where they are? While these three questions characterize more and

^{*} I thank the audiences of my courses at the 2009 LSA Linguistic Institute in Berkeley and the 2010 DGfS-CNRS Summer School on Linguistic Typology in Leipzig for helpful questions and comments on the themes underlying the present chapter. Many thanks also go to Michael Cysouw, Bernd Heine, Steven Moran, Heiko Narrog, Johanna Nichols, Sabine Stoll, Fernando Zúñiga and an anonymous reviewer for very helpful comments on an earlier draft. None of them should be held responsible of course for any remaining shortcomings or misconceptions. Part of the research summarized here was supported by Grant No. II/83 393 from the Volkswagen Foundation.

more research in typology, there has only been sporadic discussion of their theoretical foundations (briefly surveyed in the ‘further reading’ section at the end of this chapter).

The present chapter contributes to this discussion, with the aim of sharpening the contours of the emerging paradigm. Specifically, I develop one concrete approach to the ‘what’s where why’ question, called here ‘Distributional Typology’. In Distributional Typology, the *what* question is approached by the idea that the core units of language are individual structures (of any kind and granularity: constructions, relations, orderings, distinctive features, rule domains, meanings etc.) which can all be exhaustively described by large matrices of cross-linguistically applicable variables. The *where* question is explored by statistical models of how these individual structures evolve over time and across space. The *why* question finally is answered by causal theories from which these statistical models can be derived and which are themselves grounded in specific forces that are known to affect language change, from processing preferences to the historical contingencies of language contact.

I first discuss the theoretical status of generalizations in Distributional Typology as opposed to other approaches (Section 1). I then raise the *what* question: how can we find out what linguistic structures there are, and how can we compare them (Section 2)? Section 3 addresses the *where* and *why* questions by discussing types of causes that are expected to shape the distribution of linguistic structures. After introducing methods for developing statistical models of such causes in Section 4, I illustrate their use in a case study on the distribution of case marking and word order patterns (Section 5). The chapter concludes by summarizing the state of the art and identifying some pressing issues for future research (Section 6).

1 Two ways of thinking about generalizations

Most work in linguistics relies on what one may call the Pāṇinian approach:¹ seek maximally general statements, try to explain away as many counterexamples as possible by reanalyzing them as falling outside the scope of the statement, and then account for the remaining exceptions by specifying their conditions in terms that are again as general as possible. The procedure is based on the idea that the best way of capturing facts about language is in the form of categorical rather than statistical statements. This methodological commitment underlies the Neo-Grammarian principle of exceptionless sound laws as much as the standard way of analyzing grammar in most current frameworks, from Chomsky’s (1995) Minimalist Program to the accumulation of traditional insights called ‘Basic Linguistic Theory’ by Dixon (2010-2012).

For typology, the Pāṇinian approach means that the ideal cross-linguistic generalization is absolute and categorical. If there are exceptions, one will want to explain them away by reanalyzing them, or by appeal to a general condition of exception. A classical example comes from early research by Hawkins (1983). Hawkins notes the generalization

¹in reference to the Indian grammarian Pāṇini whose work from the 5th century BCE is still one of the most brilliant examples of the approach.

that prepositions entail noun-genitive order (‘house of-father’: NGen) if numerals follow nouns (‘houses three’). But there are exceptions, such as the Tibeto-Burman language Karen which has prepositions, and postposed numerals, but where genitival dependents in NPs are preposed (GenN). Under the Pāṇinian approach to typology, one needs to explain away this exception. One could try and reanalyze the Karen prepositions as being “really” something else (e.g. as postpositions displaced at the surface, or as prefixes). Or one could introduce a condition of exception. This is what Hawkins (1983) does: since Karen has SVO order, he proposed to add ‘non-SVO’ to the conditions, resulting in the revised generalization ‘ $\text{Prep} \wedge \text{non-SVO} \rightarrow (\text{NNum} \rightarrow \text{NGen})$ ’: if a language has prepositions and non-SVO order, then, if the language places numerals after head nouns, then the language places genitival dependents after head nouns.

In line with similar observations in many other areas of linguistics (cf. e.g. Bod’s chapter in this volume), the Pāṇinian approach turns out not to be the best approach to typology. Many, perhaps indeed most, interesting cross-linguistic generalizations are statistical rather than categorical. From this point of view, it would be more interesting, for example, to ignore the exceptions for a moment and study the simpler statistical generalization ‘ $\text{Prep} \rightarrow (\text{NNum} \rightarrow \text{NGen})$ ’ as a statistical trend rather than as a categorical law of language (Dryer 1998, Cysouw 2005). This move has far-reaching consequences because statistical generalizations have a completely different theoretical status than categorical generalizations. Let us explore these differences in some more detail.

Applied to infinite sets of data — such as all possible human languages — the Pāṇinian approach comes with a justification problem: how can we establish the universal validity of generalizations? Since no sample of languages can guarantee the absence of exceptions beyond the sample, a Pāṇinian generalization cannot be justified by inspecting individual samples.² Instead, as Chomsky (1965) noted long ago, the only possible justification consists in demonstrating that the generalization can be formally derived from (i.e. proven in) a restrictive theory of possible grammars. This has a crucial implication: for any such derivation to be possible, the vocabulary of the generalizations must be identical to the vocabulary of the theory. Statistical generalizations, by contrast, do not have this requirement: they need not be formulated in the vocabulary of the theory that explains them because the generalization is related to its causes only indirectly (Dryer 2006a,b): the generalization captures specific effects that can be predicted from some causal theory.

The difference between the two approaches can be illustrated by research on word order typology. Recent work by Biberauer et al. (2008, 2014) is a good example of the Pāṇinian approach: on the basis of a restrictive theory of possible grammars (which assumes universally head-initial syntax and some mechanisms of movement), the authors formally derive a categorical generalization, the “Final-Over-Final Constraint”. This constraint rules out head-final phrases dominating head-initial phrases (i.e. it outlaws structures like $*[\alpha_P [\beta_P \beta \dots] \alpha]$, where α, β, \dots stand for syntactic categories and P

²Piantadosi & Gibson (2014) have recently shown that given the number of languages we know, we normally cannot even *estimate the probability* that a generalization is without exceptions.

for ‘phrase’). At first sight, the empirical coverage of this is fairly good. Unlike earlier approaches, the Final-Over-Final Constraint not only allows for what are known as ‘harmonic’ orders (e.g. prepositions dominating head-initial NPs like in French *dans la maison de mon frère*; or postpositions dominating head-final NPs like in Hindi *mere bhāi ke makān mẽ*, lit. ‘my brother of house in’), but also for some well-known ‘disharmonic’ ones (e.g. prepositions dominating head-final NPs like English *in my brother’s house*), as opposed to head-initial NPs dominated by postpositions. But there still are some exceptions (e.g. among Surmic languages of Eastern Africa). Given the Pāṇinian approach, these are then accounted for by an extra condition on the Final-Over-Final Constraint. One of the specific conditions postulated by Biberauer et al. (2008, 2014) is that the Final-Over-Final Constraint only applies if the heads of the stacked phrases share the part-of-speech category: it is impossible for nominal postpositions to govern head-initial NPs, but possible for non-nominal postpositions to do so. The extra condition means that any counterexample can (and should) be explained away by showing that apparent postpositions in this language really are not nominal.

For statistical generalizations, the approach is very different. Here, one starts from a causal theory that is hypothesized to bring about certain effects. Dryer (1992) and, in more detail, Hawkins (1994, 2004), for example, propose that syntax can be more efficiently parsed by the human brain if less words are needed for detecting phrase structure. If there is indeed a demand for efficiency of this kind, it should cause several effects on how syntactic constructions are formed. One of these effects would be a slight preference for harmonic orders because here few words suffice to detect stacked phrases: harmonic structures like $[_{\alpha P} \alpha [_{\beta P} \beta [\dots]]]$ and $[_{\alpha P} [_{\beta P} [\dots] \beta] \alpha]$ can be fully detected on the basis of just two words, viz. the heads α and β , while a disharmonic structure like $[_{\alpha P} [_{\beta P} \beta [\dots]] \alpha]$ requires parsing of all words, e.g. the two heads plus everything that is contained in $[_{\beta P} \dots]$. In addition, because head-initial structures always allow quicker identification of phrase structures, one expects an across-the-board effect (a ‘dominance’ principle in Greenberg’s (1963) terms) that favors certain disharmonic patterns over others: disharmonic prepositions over disharmonic postpositions, and conversely, postpositional phrases embedded in head-initial NPs over prepositional phrases embedded in head-final NPs, like Biberauer et al. (2008, 2014) claim.³

Under this view, generalizations describe effects of causes. Unlike under a Pāṇinian approach, the effects need not, and indeed cannot, be formally derived from the relevant causes because they are not stated in the same vocabulary. While the effects are described in structural terms (e.g. specific structures like $[_{PP} P [_{NP} N [\dots]]]$), the causes are described in processing terms that are far more general than the specific effects that they explain. The relevant processing principles have effects on a large range of phenomena (extensively surveyed by Hawkins 2004), and they should ultimately be grounded in the neurophysiology of processing.

³ Dryer (2005a,b) reports 50 languages with head-final NPs and prepositions. To the extent that these languages allow their PPs to be embedded in NPs, the claim lacks empirical support. For a real test, however, one would need to specifically survey embedded structures, not just available phrase types.

This view of generalizations and their explanation is common practice in virtually all disciplines in which statistical methods are used, but what is the advantage of this view in linguistics? A key advantage, and presumably the reason for success across disciplines, is that we can study very specific effects of a causal theory without having to rule out all the intervening effects that one would normally expect in such an exceedingly complex phenomenon as human language.⁴ If one wants to model a specific effect, e.g. of processing on word order patterns, one can include in the same statistical model competing effects, for example effects from competing cognitive demands (e.g. from priming and analogy). In addition, one can control for effects from completely different processes, such as effects from language contact and area formation, or cultural inertia within speaker groups, or any other effects.

A complex mix of causes is typical of human languages (and most likely also of how it has evolved; cf. e.g. Fitch 2011). This situation corresponds exactly to the standard view of causal theories and testable statistical effects in most sciences. In practice, one will often want to *abstract away* from intervening effects and further complications (setting their coefficients to 0, as it were), but, as Stokhof & van Lambalgen (2011) point out, one would not want to *idealize* the situation by completely ruling out such intervening effects as a matter of principle.

Another advantage of the standard statistical approach is that specific effects can be very small. Processing effects, for example, are likely to be extremely small in general.⁵ This makes it possible for exceptions such as disharmonic word order structures to survive for millennia in languages, i.e. to be acquired and processed without any problem. A telling example of this is the persistence of prenominal relative clauses in SVO Chinese, a pattern that goes against a universal preference and is exceedingly rare worldwide (Greenberg 1963, Dryer 1992). A possible reason why processing effects tend to be small is that they are always in competition with strong alternative factors. Given the choice between harmonic and disharmonic structures (when producing sentences or when parsing them), there is an ever-so-slight preference for speakers and hearers to select a harmonic structure over a disharmonic one. But at the same time, there will be many other factors playing a role in this as well, notably the choice made by one's interlocutors and a general reliance on a system that one knows and that 'just' works, no matter how harmonic it is. As a result, we expect the processing effect to shine through only if there is a sufficient number of occasions for choice, and that there will always be exceptions.

A third, more practical than theoretical advantage of the standard statistical approach is that generalizations can be modeled using the same mathematical *lingua franca* that has proven to be extremely successful across many other disciplines. And the causes

⁴Indeed, a complex mix of interacting causes is the general motivation for adopting the 'causal-theory-and-statistical-effects' approach in most disciplines. Theories predict highly specific effects (e.g. of a particle in physics, or of a drug in chemistry), but these maybe so strongly entangled in a web of confounds that they can only be investigated by statistical modeling.

⁵Communicative needs, by contrast, tend to have much stronger effects. For example, virtually all languages distinguish formally between statements and question because the distinction is critical for successful communication. Exceptions here are extremely rare (Dryer 2005d) and presumably unstable.

behind the effects can be stated directly in the psychological or sociological theory of the domain that they are based in (e.g. in the neurophysiology of language processing). There is no need to formulate one's explanatory theory in a metalanguage that is full of notions that are unique to linguistics (such as 'merge', 'probes and goals', 'phases', 'c-command', etc., see for example Boeckx, this volume) and totally insulated from the rest of the cognitive and social sciences.

While explanations and causal theories can (and should) be formulated in the vocabulary of their respective domains (society, communication, the brain etc.), their effects, i.e. statistical generalizations, need a genuinely linguistic vocabulary, a *metalanguage* for describing such phenomena as 'relative clause', 'preposition', 'verb object order' etc. However, unlike when working with typological generalizations under a Pāṇinian approach, the choice of metalanguage here is not constrained by how one wants to explain the generalizations. Metalanguages need to allow analyses that cover the data fully (with no data left behind) and that are explicit, logically coherent and replicable. Ideally, the constructs and notions of the metalanguage have psychological reality in the sense that they are informed by what we know about how children learn languages and how languages are processed. However, as the plethora of available metalanguages ('theoretical frameworks') attests (for example in the present volume), even when one aims at fulfilling these criteria, there is still a broad choice. What, then, would a suitable choice be for formulating statistical generalizations of 'what's where why'? I turn to this issue in the following.

2 Typological metalanguages

The key challenge for typological metalanguages is that they need to be able to describe structures across languages in an empirically responsible way, i.e. without forcing languages into Procrustean beds. At the same time, if we want to have general insights into the *what*, *where* and *why* of languages, it is absolutely essential that languages are analyzed in comparable terms. The tension between faithfulness to particulars and the quest for comparable analyses has driven much debates in the field, and the pendulum has swung back and forth (see e.g. Haspelmath 2010 vs. Newmeyer 2010 for recent debate).

The problem rests on the observation that structures across languages are often similar but never really identical. For example, one can say that case marking in Hindi is of a similar type as case marking in Chintang (Sino-Tibetan, Nepal) in that both are "ergative", and many typological databases of the past would leave it at this statement. But there are many differences as well: in Hindi, the ergative is limited to certain tenses, in Chintang it is used generally; in Hindi it is used on all persons, in Chintang only on some; in Hindi, the ergative is opposed to an overt object marker under some conditions, in Chintang it is opposed to a general absolutive with zero exponence; etc. Similarly, in some sense, German main clauses can be said to have 'SVO' order like in English but there are many well-known differences; most importantly, the pre-verbal position in

German is not limited to ‘S’, and the way it is filled is mostly driven by information structure. How can we describe the similarities as well as the variation?

A principled way out of this dilemma is to take ‘similarity’ simply as what it really is: identity in some and difference in other regards. Such a situation is straightforwardly captured by large matrices where each phenomenon in each language is a row and where each ‘regard’ in which the phenomena can be the same or different is a variable (a.k.a. ‘parameter’, ‘feature’ or ‘character’). I call this approach ‘Multivariate Typology’ in Bickel (2010b, 2011b) and it is close in spirit to what Levinson & Evans (2010) mean by a ‘Low-Level Feature Metalanguage’.⁶ The resulting matrices are in principle similar in design and intent to the genomic databases that have revolutionized biology. The basic idea of Multivariate Typology is to develop variables that are maximally fine-grained (and become ever more fine-grained the deeper our analyses go).

Table 1 gives a simplified example of such a matrix, based on current research on case marking (Bickel 2011a, Witzlack-Makarevich 2011). The matrix contains sample entries of case markers in Chintang (ISO693.3: *ctn*) and Hindi (*hin*). The ID in the second column links to other matrices, for example, one matrix that describes the phonological and morphological properties of case markers, including specific representations of the shape of markers (1372: *-ŋa*, 92: *-ne*) and more general issues such as whether markers select for specific hosts, whether they are phonologically free or bound, whether they undergo lexical allomorphy etc. The other columns in Table 1 are briefly explained in Table 2.

The typology tries to factor out as many aspects of case markers as are descriptively justified across languages and of potential interest for statistical generalizations. From a descriptive point of view, the typology can be seen as a condensed version of the information that one would want to include in a reference grammar. For formulating generalizations and explore ‘what’s where why’, one can derive various typologies of interest. For example, by comparing the sets of generalized semantic roles (‘S’, ‘A’, etc.) that are covered by case markers, one can automatically compute for each language whether, and if so, under which conditions, the language marks a distinction between the arguments of transitive clauses. Such a typology is of interest to various processing hypotheses, one of which I will come back to in Section 5. Applying more complex algorithms to the typology in Table 1, one can also automatically compute statements of role alignment (‘accusative’, ‘ergative’ etc.), relative to referential categories (e.g. one per person), verb categories, and clause types, and then test hypotheses on their areal and universal distribution with great precision (see Bickel et al. in press-*a*, *-b*, for applications). The variables in Table 1 can also of course be linked to any other set of variables, e.g. to variables that capture the semantics of predicate classes. The matrices

⁶ Another similarly-minded approach is Corbett’s (2005) ‘Canonical Typology’, which shares an interest in working with multiple parameters of comparison at once. Canonical Typology is special in assuming canons as ideal starting points where Multivariate Typology aims at measuring similarities without *a priori* weighting of specific phenomena (while still allowing ad-hoc weighting in order to test specific hypotheses). Multivariate Typology was first sketched in Bickel (2007); a proof-of-concept study is Bickel (2010b).

ISO639.3	ID	Role	PoS	co.Role	co.PoS	PredCat	Clause	Predicate Class
ctn	1327	A _d	non-excl	ANY	ANY	ANY	ANY	Primary object verbs: some verbs denoting covering events, events of destructive impact like throwing, kicking, hitting, or cutting
ctn	1327	A _d	N	ANY	ANY	ANY	ANY	Primary object verbs ...
ctn	1327	A _d	non-excl	ANY	ANY	ANY	ANY	Double object verbs: physical and mental transfer events (translated as 'send, bring, take, move to, give, pass to, infect, feed, tell, ask for, show' etc.), also verbs like yukt- 'to keep for someone', which represent a kind of 'intended transfer'; verbs of covering ('cover, bury, pour, throw, spray at, soil, stain,' etc.)
ctn	1327	A _d	N	ANY	ANY	ANY	ANY	Double object verbs ... (as defined above)
ctn	1327	A _d	non-excl	ANY	ANY	ANY	ANY	the default ditransitive predicate class
ctn	1327	A _d	N	ANY	ANY	ANY	ANY	the default ditransitive predicate class
ctn	1327	A	non-excl	ANY	ANY	ANY	ANY	the default transitive predicate class
ctn	1327	A	N	ANY	ANY	ANY	ANY	the default transitive predicate class
ctn	1327	T	non-excl	ANY	ANY	ANY	ANY	Primary object verbs ... (as defined above)
ctn	1327	T	N	ANY	ANY	ANY	ANY	Primary object verbs ... (as defined above)
ctn	1334	P	N	ANY	ANY	ANY	ANY	the default transitive predicate class
ctn	1334	A	excl	ANY	ANY	ANY	ANY	the default transitive predicate class
hin	92	A _d	ANY	ANY	ANY	PP-hin	main	the default ditransitive predicate class
hin	92	A	ANY	ANY	ANY	PP-hin	main	predicates with ERG depending on 'conscious choice' or volitionality (alternation possible only in perfective): samajh 'understand, suppose', bhul 'forget', jan 'give birth (to)', phāḍ 'leap over', bak 'to talk nonsense', har 'lose, be defeated' (Butt 2001: 127)
hin	92	A	ANY	ANY	ANY	PP-hin	main	the default transitive predicate class
hin	92	S	ANY	ANY	ANY	PP-hin	main	predicates with ERG/NOM-Sintr/Atr depending on 'conscious choice' or volitionality (alternation possible only in perfective): intr. verbs: bhōk 'bark', jhāk 'peep, look into/through', khās 'cough', chīk 'sneeze', muskara 'smile', thuk 'spit', mut 'urinate', hag 'defecate', naha 'bathe', ro 'cry', hās 'laugh', so 'sleep' (Butt 2001: 127)

Table 1: Simplified sample entries for the ergative case markers in Chintang and Hindi in a Multivariate Typology (Witzlack-Makarevich 2011). 'ANY' means that there are no constraints.

<i>Variable</i>	<i>Description</i>
Role	Generalized semantic role, based on the number of arguments and lexical entailment tests. <i>Possible values</i> : with 1 argument: S; with 2 arguments: A, P; with 3 arguments: A _d , T, G.
PoS	Parts of speech and related categories. In the real database, abbreviations like “non-excl” (for all person pronouns except first person dual and plural exclusive) are fully resolved into all person pronoun categories that the language has. <i>Possible values</i> : open, mostly cross-linguistically applicable categories; some language-specific residues like pronouns with gender specification (e.g. Russian neuter pronouns)
co.Role	Co-argument role: in some languages (such as Umatilla Sahaptin: Rigsby & Rude 1996), role marking of one argument depends on the co-presence of a certain other role, sometimes with a certain Part of Speech (e.g. ergative case marking on A only if there is a first or second person P argument). <i>Possible values</i> : (see under ‘Role’ above)
co.PoS	(like PoS but for co-arguments)
PredCat	Categories that constrain the assignment of a role marker. <i>Possible values</i> : mostly language-specific (e.g. ‘PP-hin’ for Past Participle in Hindi) but in the real database this is to some extent decomposed into formal (e.g. periphrastic vs. synthetic) and semantic (e.g. perfective vs. imperfective) components
Clause	Constraints on clause types, e.g. main vs. dependent clauses. For example, in Table 1, the Hindi ergative case marker is specified as occurring only in main clauses. <i>Possible values</i> : largely cross-linguistically recurrent types, developed further in Bickel (2010b)
Predicate Class	Lexical predicate class, with an abbreviated description in Table 1. <i>Possible values</i> : language-specific, but in the real database the classes are given as sets of individual predicates and are annotated for cross-linguistically recurrent meaning components (such as ‘experience’, ‘perception’, ‘obligation’ etc.)

Table 2: Some variables in a Multivariate Typology of case markers (for details, see Bickel 2011a, Witzlack-Makarevich 2011)

that result from this can then be mined for recurrent types or prototypes, for universal or area-specific clusters, or for potential correlations between variables, applying any kind of suitable data mining algorithm (for some examples, see Croft & Poole 2008, Cysouw et al. 2008, Bickel 2010b, Donohue et al. 2011 etc.).

No Multivariate Typology is ever complete: in the example table, notions like ‘main’ vs. ‘dependent clause’ are extremely abstract and gloss over much (too much!) variation across and within languages. In this sense, Multivariate Typology is more a research program than a theory: the goal is to decompose every single descriptive notion into evermore fine-grained variables and thereby achieve at the same time greater precision in description and richer opportunities for detecting generalizations. This goal is best served if the typology is maximally general and contains as few language-particular notions as possible. This is of course a tall order, but it is part and parcel of any attempt to understand the phenomenon of human language in a way that goes beyond cataloguing particulars. Fortunately, there has been and continues to be considerable progress, as testified by the dual increase of typologically well-informed descriptive grammars and

descriptively ever more precise typological studies. The biggest challenge here is still semantics. For example, for full comparability the variables that are hidden behind ‘Predicate class’ in Table 1 need detailed expansion into universal semantic features, perhaps along the ways proposed by Goddard (this volume). A promising alternative are denotational values that can be used as stimuli in elicitation (as is traditionally done for example in color term typology). There is no constraint on the kind of variables one needs. Variables can be about features and denotations as much as about more abstract variables, specifying such structures as sets (e.g. of predicates, of roles), domains (e.g. of rule application in phonology) or orderings (e.g. before vs. after).

However, there is a natural bottom end to the enterprise: arbitrary sound/meaning links, i.e. Saussurian signs like Chintang *-ŋa* ‘ergative’. These are language-specific by definition and no matter how well we capture the sound and the meaning by typological variables, Saussurian signs are individuals by definition, and as such, they are not genuinely repeatable across languages like, say, a specific role set. While Saussurian signs do not support typological generalizations, they still play an important role for typology: they are the key data for positing genealogical relatedness, an issue to which I will return later. Before this, however, I wish to highlight one further theoretical aspect of Multivariate Typology.

Multivariate Typology departs from traditional ways of formulating metalanguages for linguistic analysis. In traditional metalanguages, saying that a particular phenomenon instantiates a given universal category is expected to capture more than one specific property. For example, saying that some case marker instantiates a ‘nominative’ lets us expect a whole range of properties: at least a set of roles covered (S and A, and not object or ‘goal’ arguments of any kind) and certain syntactic privileges associated with this set (such as triggering agreement if there is agreement), etc. A consequence of this is that one can speak of ‘diagnostics’ or ‘tests’: whether a specific phenomenon instantiates a given universal category can be ‘diagnosed’ via the range of properties that are expected from the category (‘if an NP marked by X triggers S and A agreement, X is a nominative case’). As shown by Dryer (2006b), such a view is warranted only if metalanguages are expected to be explanatory, e.g. to be able to say that a certain set of roles is syntactically privileged *because* it represents the ‘nominative’ set (as this is how we expect nominatives to behave). As argued above, Distributional Typology distinguishes between metalanguages and explanations. If there are linkages between variables (e.g. between case marking covering S and A and agreement with S and A), these linkages are *a posteriori*, not *a priori*. If variables do correlate empirically, this represents a statistical (not categorical) generalization. And if it is a genuine generalization, we expect it to be the result of some causal theory. This brings us to the next question: what is the nature of the causes that explain distributions?

3 Types of causes

Virtually all known causes of typological distributions are historical processes: they are forces that push language change in one way or the other, i.e. they create certain

systematic biases (preferences, trends) in diachronic development. There are basically two classes of theories for this, functional and event-based theories.

- **Functional theories:** these theories cover the large class of cognitive/physiological and social/communicative principles that can be hypothesized to bias the development of specific structures in certain ways, e.g. processing principles that favor word order harmony across phrase types. Functional theories predict that language change tends to go in a specific direction because languages follow preferred pathways of change vs. non-change (e.g. grammaticalization channels) or that they are subject to preferred selections of variants (e.g. expansion or preservation of some constructions at the expense of others). Functional theories often have a global scope and propose universal principles with worldwide effects, but this need not be so: one could as well have a functional theory of local (population-specific) principles and their effects. A plausible theory of this kind for example would propose that certain patterns in the language of kinship are caused by corresponding social practices (cf. Evans 2003 for discussion of relevant examples). Other examples come from Trudgill's (2011) theory that different sociological constellations (such as contact vs. isolation, dense vs. loose social networks, small vs. large community size) explain to some extent the rise and fall of complexities in grammar.
- **Event-based theories:** when languages are in contact, speakers often copy structures from other languages or develop similar structures that mimic patterns in other languages. Event-based theories account for those processes of copying and replication that are not grounded in how well structures fit with the way our brain or communication works, but that instead result from whatever happens to be popular and *en vogue* in a given situation during a given time. Thus, event-based theories are always grounded in the contingencies of the contact history of languages, and they are therefore tied to specific geographical areas and specific (series of) historical events in them, e.g. the spread of HAVE-perfects in Europe in the transition period between antiquity and the middle ages (Heine & Kuteva 2006). In order to avoid circularity, it is critical that event-based theories are fully based on findings in other disciplines, from history to archeology and population genetics (Bickel & Nichols 2006). Note that unlike functional theories, event-based theories do not generalize beyond the accidents of history. Event-based theories only predict that languages tend to develop in different directions inside than outside a hypothesized area because the relevant structures were replicated by speakers just for their popularity at the time, and not for any functional reason.

Event-based theories capture most cases of language contact processes, but not all of them. Some processes during contact are clearly also affected by what is best captured by functional theories: for example, structures that are easier to learn for adults can be generally expected to spread in contact more easily than more difficult structures. Perhaps there are even whole parts of grammar that are relatively immune against borrowing

and change (i.e. that are relatively ‘stable’), because they are too tightly integrated in the way the grammar is processed. Also, given certain types of contact situations (such as substrate situations), the diffusion of specific structures across languages may be affected by the complexities of second language learning and bilingual processing. In all these cases, areal diffusion, or the systematic absence thereof, reflects functional factors that have universal validity and are not tied to a specific event (or to a specific series of events) in history: with a certain probability, the diffusion is expected to apply or not apply between all languages in contact, or between all languages in contact that meet certain structural or sociological conditions.

This difference between functional and event-based factors in language contact has an important methodological consequence: functional factors in contact can be estimated as universal probabilities of horizontal transmission (as done for example by ?), but this is not possible for event-based factors because they are triggered by individual contingencies of history rather than universal principles of borrowability. This also means that there is no statistical shortcut around detailed research on history and prehistory.

Functional theories can be about any kind of factor that plausibly determines how languages develop. While in most typological work the relevant theories are located externally to grammar, this does not have to be so. It is perfectly possible and compatible with the overall approach of Distributional Typology to entertain ideas like the proposal of recursion as a genetically given dimension specific to grammar (Hauser et al. 2002, Chomsky 2010). Unlike in the Chomskyan tradition, however, in Distributional Typology, the interest of such a theory would not be in constraining our metalanguage (since, as argued in Section 1, metalanguages have no stakes in explanation). Instead, reframed in terms of a functional theory, the hypothesis of interest would be this: the development of individual languages over time is expected to fit better with the genetics of our brain (and the cognitive practices associated with this, e.g. in visual pattern recognition: Stobbe et al. 2012), if languages employ hierarchical and recursive syntax rather than simple juxtaposition (e.g. *I set up the computer that just arrived* rather than *The computer just arrived. I set it up*). In other words, one would expect our genetic heritage to cause statistical effects on diachronic biases, e.g. it should be more likely, *ceteris paribus*, for languages to develop complex embedding than losing it (as indeed is often claimed for Indo-European, cf. Viti 2012 for a recent review of the evidence).⁷ Such a genetically-based theory would be entirely parallel to such theories as the recent proposal that a certain haplogroup distribution in a population favors the development and maintenance of tonemic distinctions (Dediu & Ladd 2007, Dediu 2011) in the same population — if only now at a global (species-wide) rather than local (population-specific) level.

From this perspective, the traditional distinction between language-external and language-internal causes for linguistic change and evolution may turn out to be of little

⁷ At present, there are no sufficiently large databases available for testing such claims on a worldwide scale, but see e.g. Karlsson (2009) and Mithun (2010) for suggestions about the variability of recursion in syntax and its limits.

interest in the end — a conclusion that coincides with what Fitch (2011) arrives at from a biological perspective. What is of much greater concern from a linguistic point of view is to determine to what extent functional factors (internal or external) are independent of contact events in history (Dryer 1989): a functional theory based on universal patterns, such as the way our brain works, is supported only if its predicted effects on structure (e.g. a preferred type of phrase structure) hold independently of historical contact areas. For example, without detailed research, it is unclear whether the rarity of verb-initial word order is best explained by a universal theory (be it in terms of formal grammar design as in Baker (2001) or be it in terms of processing as in Hawkins (2004)) or rather by long-lasting series of contact events around the Pacific Rim, in the eastern Rift Valley and across insular northwest Europe (i.e. in the areas with increased frequencies of verb-initial order). A key task for any approach to ‘what’s where why’ is precisely to develop models that allow to tell such factors apart.

4 Methodology

How can we formulate statistical models that describe the ways in which current distributions of structures are brought about by specific factors? Traditionally, typologists have sought to solve the problem by assuming that we can identify the relevant factors by ‘weeding out’ as it were genealogical relatedness and then analyze the remaining statistical distribution of structures. The assumption is that if languages are genealogically related, it is always possible that they share structures not because of the relevant factors, but because they descend from the same proto-language (e.g. Spanish and Italian would have subject agreement simply because Latin had it). Therefore, all that is needed from this traditional point of view is to exclude this possibility, either through strategic sampling (cf. Bakker 2011 for a review of the tradition here) or by including genealogical relatedness as a confounding variable in statistical models (e.g. Bickel et al. 2009, Jaeger et al. 2011).

However, when related languages share structures, this itself can provide key evidence for functional or event-based factors. Being related only means that languages arose through splits away from a proto-language. If a specific structure persists during such splits, this could mean that the structure was favored by some factor (e.g. of processing or of a series of contact events in a linguistic area) just as well as it can mean that the structure survived because of mere chance – just like a change in a variable can reflect both the effects of some factor or of chance (cf. Maslova 2000, Nichols 2003, Bickel 2013). It is essential for a statistical model to be able to identify all these possibilities.

Models that meet this requirement can be developed based on the following idea: given the current distribution within a set of genealogically related languages (a language family), we can statistically estimate the most likely history of change and non-change that has led to the current distribution, i.e. we can estimate diachronic developments, and then examine to what extent these are the products of chance vs. to what extent they are systematically influenced by functional or event-based factors. Such estimates are of course far from trivial, and since we hardly ever know the true history of a family (only

very few proto-languages are attested through documents), any resulting estimate cannot be directly calibrated against reality.⁸ Nevertheless, there are now various concrete proposals available that implement the basic idea of statistically estimating diachronic trends from synchronic data: proposals by Maslova (Maslova 2004, Maslova & Nikitina 2007), by Dunn and colleagues (Dunn et al. 2011), and by myself (Bickel 2011c, 2013). In the following I limit the discussion to my own proposal, the ‘Family Bias Method’, because (as far as I am aware) it is the only available method that generalizes to data from small families (with only a handful of members) and isolates. Furthermore, unlike Maslova’s method, the Family Bias Method allows models for causal theories of any complexity, with any number and kind of factors, including scalar and multinomial factors. Finally, unlike the method adopted by Dunn and colleagues, the Family Bias Method is equally applicable to typological data that are variable or constant in a family. Furthermore, the Family Bias Method does not require that we necessarily assume tree structures instead of wave models, while it can make use of tree structure information if it is available.⁹

In the Family Bias Method, each family is evaluated as to whether its daughter languages show a bias towards the same structure (or set of structures, or a correlation between structures, etc.), as revealed by a suitable statistical test (or a Bayesian estimator). If there is such a bias (a statistically significant number of daughter languages have the same structure, e.g. OV order), this means that daughter languages have preferentially innovated in the same direction, or they kept what was already in the proto-language. Either way, a bias suggests that — for whatever reason — there was a systematic preference in the development of the given family. The absence of a bias suggests random fluctuation in development. Note that at this point it does not matter what causes are responsible for biases or their absence: they can be chance effects that happened to affect a specific family for some time, or they may result from event-based diffusion in an area, from universal principles, or any other cause. Also, it does not matter what the structures really were in the proto-language: for estimating a bias, we don’t need to know because the same bias can result from multiple innovation as well as from systematic preservation (or from early innovation followed by systematic preservation).

The presence of biases of this kind can be determined straightforwardly in families with enough representatives — say, at least half a dozen. In these cases, an appropriate test is for example a binomial (or multinomial) test, but other procedures are possible as well. But what about smaller families, or families with just one member, i.e. isolates? Here, biases can be estimated using extrapolation algorithms: for this, we can use the information on biases in large families in order to estimate the biases that are likely to have been behind the attested structures in small families: if, say, 60% of large families

⁸Like in other disciplines with this problem, the obvious alternative is to evaluate methods by computer simulations of language change, but research here is only beginning.

⁹The Family Bias Method is implemented in an R package (R Development Core Team 2014) developed by Taras Zakharko and available for download at <https://github.com/IVS-UZH/familybias>. For detailed exposition of the method and justification of its assumptions, see Bickel (2013).

are biased towards some specific structure (e.g. biased towards OV order, or towards VO order) rather than balanced between structures (i.e. with about as many OV as VO daughters), we estimate a .6 probability that the known members of small families come from larger unknown families with a bias as well (in whatever direction), as opposed to families without any bias.¹⁰ Some of the known members will be representative of the bias in the unknown larger family, and so we can take their structural choice (e.g. VO order) to reflect the bias (picking the majority value, or making a random choice in the case of ties). But some known members will happen to be deviates, e.g. the odd guy out that developed an OV pattern although the family as a whole is biased towards VO (like Tunen in Bantu: Mous 1997). The probability of being representative can be estimated from the strength of the bias in large families: e.g. if among biased large families, biases tend to be very strong (e.g. on average covering over 90% of members), we can estimate a high probability that the known members of small biased families are representative of the larger unknown family from which they derive; then, the probability of being the odd guy(s) out is much lower.

In summary, using the probabilities of bias and of representativeness based on large families, we can estimate how many of the small families come from larger biased as opposed to unbiased families, and if they come from biased families, we can estimate whether the known members reflect the respective biases of their families or deviate from them. These extrapolation estimates introduce error but do so randomly along a normal distribution. Therefore, we get a fairly reliable estimate of family biases if we extrapolate many times (say a few thousand times) and then compute the average of this.

The resulting estimates on family biases provide appropriate data for exploring the various causes that can affect diachrony, and that are therefore responsible for ‘what’s where why’. A hypothesized cause is supported to the extent that it accounts well for the estimated biases, given all possible interactions with effects of other causes. When families show no biases, by contrast, there are no preferences, and proto-languages seem to have developed randomly. A straightforward way of exploring these possibilities systematically and in one go is to explore the extent to which estimated family biases are statistically associated with the effects of relevant causes. A suitable framework for this is the family of generalized linear models because this statistical technique allows one to directly estimate the relative contribution of each effect. I illustrate this by a case study in the following.

5 A worked example: case marking, word order and the Eurasian spreads

Many typologists have hypothesized that a verb-final clause structure favors the development or persistence of markers that distinguish the roles of argumental NPs, i.e. case markers or adpositions (Greenberg 1963, Nichols 1992, Siewierska 1996, Dryer 2002).

¹⁰Technically, a more appropriate estimator is Laplace’s Rule of Succession which avoids unrealistic probabilities of 1 and 0 in extreme cases. See Bickel (2013).

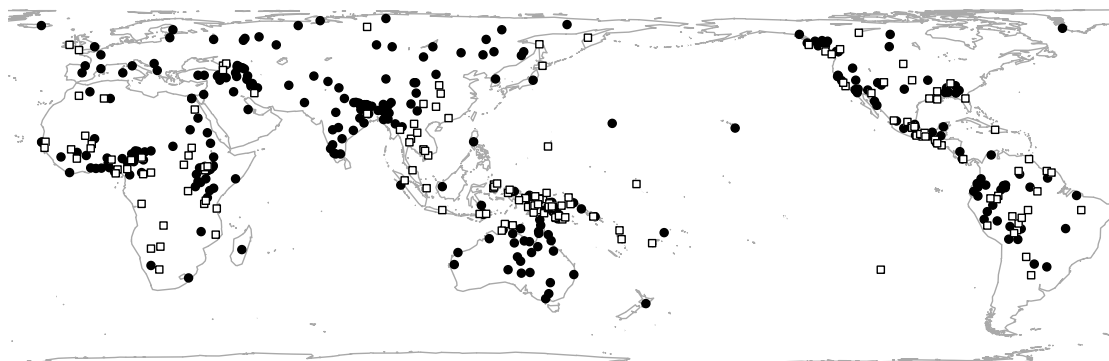
Hawkins (2004) presents a functional theory of what would plausibly cause such an effect: under this theory, the presence of role-distinguishing markers on NPs allows quicker identification of argument structure. For verb-final clauses this has the advantage that the parser does not have to wait until it encounters the predicate that makes explicit the argument structure. Although Hawkins does not discuss this, it seems reasonable to assume that what matters is distinctions of any semantic role: agents vs. patients just as well as experiencer vs. stimuli. Among all role pairs, the ones with human referents in both roles arguably play the most important role, since it is here that confusions are most likely, typically with frequent predicates like ‘see’, ‘hear’, ‘tell’, ‘meet’, ‘follow’, etc. In many languages, constellations of two human referents with the relevant predicates have special, ‘non-canonical’ case marking. Therefore, the theory is best tested against a multivariate typology that includes the full range of possibilities and is not limited to what is often thought of as canonical agent-patient pairs (in e.g. verbs of hitting or breaking).

Thus, for current purposes, a language counts as having case marking if it formally differentiates between two argumental NPs of at least one kind (e.g. only first and second person pronouns) in at least some bivalent predicates (e.g. perhaps only in some experiencer predicates with an oblique experiencer). A sufficiently rich database is the one illustrated in Table 1 above. For testing the theory, this database is merged with data on verb-final vs. non-verb-final order from AUTOTYP (?), enriched by Dryer’s (2005c) data. The database contains 489 languages distributed over 29 families with at least 5 members and 120 smaller families.¹¹ For the processing hypothesis, it is reasonable to expect the strongest effects in the most common clause types (e.g. in main rather than dependent clauses), and this happens to be what the database is limited to.

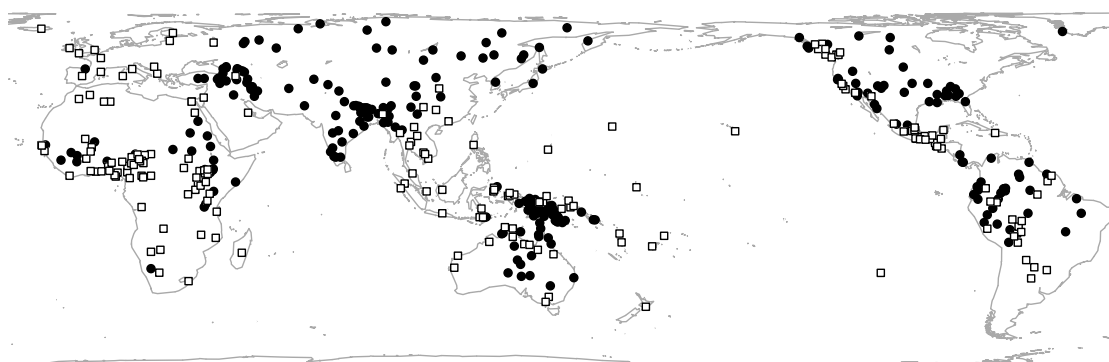
Maps 1 and 2 show the worldwide distribution of the two variables. Comparison of the maps indicates that both variables are subject to areal diffusion patterns, in line with many suggestions in the literature (Dryer 1989, Siewierska 1996, Dryer 2000, 2005c, Bickel & Nichols 2009): Eurasia is known to favor case marking whereas Africa is known to disfavor it. Southeast Asia and Europe are known to favor VO order while the rest of Eurasia is known to strongly favor OV order. This raises the possibility that the association of case markers and verb-final order is an artifact of a joint spread of these phenomena in some parts of the world during some period of time, i.e. that the current distribution is best explained by an event-based rather than a functional theory.

Since case markers predominate all over Eurasia, the most pressing event-based theory to consider is Jakobson’s (1931) classical proposal that Eurasia is a linguistic area as a whole. There is quite a bit of evidence for intense contact and many repeated language shifts in the wake of conquests throughout the known history of the area (Nichols 1992, 1998). Genetic findings by Rootsi et al. (2007) suggest that similar processes of male-dominated diffusions have had much earlier origins (dating back to over 14-19k years

¹¹Merging the word order data of the two databases is justified because for all 270 languages coded in both databases, the coding is identical. The full dataset used in the present analyses is available for download at <http://www.autotyp.uzh.ch>.



Map 1: Presence (filled circles) vs. absence (empty squares) of case markers: 83% of the languages in Eurasia vs. 60% elsewhere have case markers for distinguishing A and P in at least some NPs and with at least some predicates



Map 2: Final (filled circles) vs. non-final (empty squares) verb positions: 74% of the languages in Eurasia vs. 53% elsewhere put verbs final in the most common declarative main clauses.

ago) and with a much wider range (linking Southeast Asia in counterclockwise direction with Northern Europe). Therefore, it is entirely possible that the current distribution of case markers and verb-final orders reflects ancient spreads. Southeast Asia (and to a more limited extent also Europe) now departs from this, but these could plausibly reflect later developments (see Enfield 2005 for Southeast Asia and Haspelmath 1998 for Europe). If the Eurasian spreads are significant, we expect them to leave statistical signals despite these later developments (e.g. in the form of case markers and remnant verb-final structures that survive in most Tibeto-Burman in Southeast Asia, resisting the overall profile of the area).

What we have, then, is two competing theories of diachronic developments: a functional theory that suggests that processing advantages lead to a coupling of case marking and verb-final order over time and an event-based theory that suggests that this coupling was caused by accidentally joint diffusion of the two features. The two theories can be tested by estimating family biases for developing case markers under the two conditions of a family having (a) verb-final vs. other order and (b) of a family being placed inside

vs. outside Eurasia (the latter defined for current purposes as in Nichols & Bickel 2009). The question is which of these conditions (or both in interaction) best predicts any biases within families to maintain and/or innovate case markers.

There is a small problem for such a test design, though: some families are split with respect to the relevant features, e.g. 17 families are split between final and non-final order (e.g. Indo-European) and two families span beyond Eurasia (Austronesian and Afroasiatic). To solve the problem, we can search for the lowest taxon that is not split by these factors and determine family biases within the resulting taxa. In Indo-European, for example, such taxa can be found at major branch levels. When no lowest non-split taxon can be found or the genealogical database I use (Nichols & Bickel 2009) does not contain enough subgrouping information (which is the case in 8 out of the 19 split families), I set up pseudo-taxa based on the relevant conditions, e.g. a verb-final pseudo-group and a non-verb-final pseudo-group within Arawakan. These pseudo-groups were invariably small (containing between 1 and 4 members), and so they entered the analyses only during extrapolations to small families.¹²

Combining the analysis of large families with averaged extrapolations to isolates and small families (including pseudo-groups), we obtain the frequency distribution displayed in Figure 1.

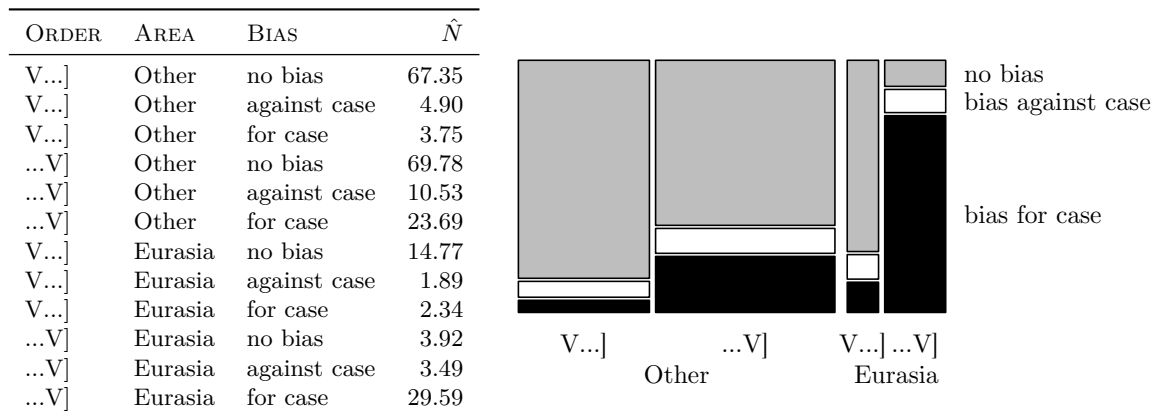


Figure 1: The distribution of estimated biases in developing or maintaining case markers within genealogical taxa across verb-final vs. other orders and Eurasia vs. other areas. The plot on the right visualizes the relative proportions of the counts in the table on the left, using the ‘mosaic’ technique provided by Meyer et al. (2006).

The relative counts of families estimated as lacking a bias (grey tiles in the plot) suggest that there are many families without a significant bias towards or against case markers, i.e. case markers come and go regardless of word order or area conditions. This is especially true outside Eurasia: families seem to be more diverse with regard to case-marking here than in Eurasia; within Eurasia, families are more diverse in this regard in non-word-final than in final word order groups. These findings have their own interest

¹²For a general discussion of pseudo-groups, including large ones, see Bickel (2013).

and invite further research on the sociological reasons for diversification in families (for a start, cf. Nichols 1998 on Eurasia and Trudgill 2011 for a theoretical framework).

For now let us focus on the effects that functional and event-based theories are expected to have on biases towards or against case marking. For this, families without biases provide no evidence: unless we know the proto-language (which is almost never the case), structural diversity can arise from some languages changing towards a specific structure or away from it. Concentrating on families with biases, Figure 1 suggests that biases towards rather than against case marking (black vs. white tiles) are found more often in Eurasia than elsewhere (see the difference between the ‘Other’ vs. the ‘Eurasia’ two-column panels in the figure) and more often if the family is verb-final than not (cf. the ‘V...’ vs. ‘...V’ columns inside each area panel). This fits with the predictions of both the functional and the event-based theory, but could the effects be the product of chance processes?

To find out, the frequencies of families with biases towards vs. against case markers can be studied in a generalized linear model (GLM). For this, I assume, with Cysouw (2010), that linguistic diachrony reflects underlying Poisson processes; a suitable GLM with this assumption is known as the log-linear model. These models allow one to systematically explore whether specific interactions of columns (value combinations) in a table like the one in Figure 1 are needed in order to account for the differences in the counts: for example, without an interaction between the area and the bias columns (formally, $\text{AREA} \times \text{BIAS}$, combining values like ‘Eurasia’ & ‘for case’, ‘Eurasia’ & ‘against case’ etc.), one would expect that the ratio of ‘for case’ vs. ‘against case’ counts within each word order type is roughly the same in Eurasia and elsewhere, and that any observed difference falls within what one expects to result from random (Poisson) processes alone. Intuitively, this does not seem to be the case in the data at hand: the differences are unlikely to result from chance (e.g. within verb-final families, the ratio is $29.59/3.49 = 8.48$ in Eurasia vs. $23.69/10.53 = 2.25$ elsewhere), and one could not successfully model the observed data without the $\text{AREA} \times \text{BIAS}$ interaction.

A standard procedure to test such interactions formally rather intuitively is known as an *analysis of deviance*. For the present data, such an analysis suggests that the three-way interaction $\text{BIAS} \times \text{AREA} \times \text{ORDER}$ and the two-way association $\text{AREA} \times \text{ORDER}$ are not needed, but that the two two-way associations of interest, $\text{BIAS} \times \text{ORDER}$ and $\text{BIAS} \times \text{AREA}$, make both statistically significant contributions.¹³

This means that biases towards case are associated with both verb-final word order and location in Eurasia. Since there is no three-way interaction, these two-way interactions are independent of each other, i.e. the underlying causes in processing and areal diffusion appear to work independently of each other. This finding converges with a

¹³The analysis of deviance was performed with likelihood ratio χ^2 tests comparing the deviance (‘lack of fit’) of successively simpler models, i.e. models with vs. models without the relevant interactions (using the `anova.glm` function in R). The results are for $\text{BIAS} \times \text{AREA} \times \text{ORDER}$: $\chi^2 = .551, p = .45$; for $\text{AREA} \times \text{ORDER}$: $\chi^2 = .554, p = .45$; for $\text{BIAS} \times \text{AREA}$: $\chi^2 = 4.542, p = .034$; and for $\text{BIAS} \times \text{ORDER}$: $\chi^2 = 4.903, p = .027$. These same conclusions can be reached by stepwise model selection using Akaike’s *An Information Criterion* (AIC), implemented as `stepAIC` (Venables & Ripley 2002).

recent study by Sinnemäki (2010) that uses the Family Bias Method to explore the association of zero role marking (i.e. absence of any case or agreement) with both word order patterns (here, SVO order) and areal diffusion. Like in the present study, both types of causes have a significant effect but appear to operate independently of each other.

6 Conclusions and outlook

The case study above has shown how we can find out about trends in diachrony from densely sampling families. The method proposed therefore fits with a fundamental tenet of Distributional Typology (cf. Section 3): that the distribution of structures is always the product of history. The diachronic approach also solves a troubling question of representativeness: to what extent can a sample of the languages that happen to be spoken now (and happen to be investigated) ever tell us anything about human language in general? The problem is that the current range of languages may not (and is indeed unlikely to) instantiate everything that is humanly possible because languages arise and die for reasons that have nothing to do with their structures. The Family Bias Method and related methods, however, do not aim at specifying ranges of possible structures. Instead, what the methods give evidence of is specific effects of causal mechanisms that operate in diachrony. The causal mechanisms are in principle expected to leave signals in all families (with the relevant structural properties and geographical locations), regardless of the range of families that happen to exist at any given time or that one happens to sample (Levinson et al. 2011). This makes the findings independent of whether or not our samples are representative of the entire universe of human language.

However, many causes of interest are very weak. Indeed, causes that are based on processing preferences are even expected to leave only relatively small signals in the data since languages that violate functional theories can still be acquired and processed over generations. Furthermore, effects might be totally blurred and even reversed by known and unknown counteracting causes as well as by chance processes that can affect individual linguistic eco-systems. Therefore, in order to detect any statistical effects, one typically needs to cast the net very wide. This means that one needs to sample as many families as possible and sample these densely (breaking thereby with traditional sampling strategies that were concerned with picking only one or only few languages per family). Also, in order to identify critical interactions of functional and event-based effects, it is essential to sample families across all geographical regions that are known to have undergone extensive spreads.

This leaves us with an important practical call: in order to gain further insights into the *what*, *where* and *why* of language we need much larger databases. And to the extent that causal theories become ever more precise, the more they bear on very specific structures (e.g. on case marking when all referents are human and not just any case marking). This requires databases to reach higher descriptive precision. In the long run, it won't do to gloss over relevant distinctions and set up gross types like "ergative language" or "SVO language". In this regard, the needs of large-scale typological work

entirely coincide with the needs of fieldwork-based descriptive research and, as argued in Section 2, these needs are best served by developing fine-grained Multivariate Typologies.

In addition, there is an urgent need for more sophisticated statistical methods. What I have presented in this chapter is what is currently available. The methods are still a bit crude, e.g. by relying on statistical tests of whether or not a family is biased rather than on estimates of the degree to which it is biased. Distributional Typology urgently needs intensive research here, but what has become clear is that it won't do to blindly copy methods from bio-informatics or other fields where processes over time are modelled. The diachronic dynamics of language have special properties that need to be accounted for — above all the fact noted in Section 4 that the absence of change can be an important signal of systematic causes that constrain diachrony.

However, other aspects of the overall orientation of Distributional Typology are far more important than these relatively technical issues. First, because of its Pāṇinian heritage, typological research has often concentrated more on formulating generalizations than explicit causal theories that would explain them. Once generalizations are statistical, however, they are always only as interesting as the causal theory behind them. In response, much more effort is needed now to develop richly articulated theories. For functional theories, this means extensive cooperation with psychologists and geneticists on the one hand and with anthropologists and sociologists on the other hand. For event-based theories, it is imperative that theories be deeply grounded in non-linguistic evidence, and this again requires detailed cooperation with other historically- and geographically-oriented disciplines. A statistical signal of some features in an area can be entirely spurious unless it is shown to be plausibly caused by concrete historical events.

A second issue relates to the use of statistical methods. Distributional Typology is not the only approach that breaks with the Pāṇinian tradition of categorical analysis. Indeed, there is a trend in this direction sweeping across all of linguistics. There is mounting evidence that even our linguistic “competence”, impressively manifested through our ability to judge the grammaticality of any unheard utterance, is ultimately probabilistic in nature, largely predictable from previous experience (Bresnan 2007, Bresnan & Ford 2010; Bod, this volume). While Distributional Typology has so far mostly targeted discrete values on nominal variables (and this is also what I focused on in this chapter), one of the most exciting challenges is now to target directly probabilistic patterns and base typology on corpora rather than reference grammars. There is indeed a rich, but so far largely untapped, range of probabilistic patterns that are likely to be driven by functional and event-based theories. For example, Bickel (2003) and Stoll & Bickel (2009) explore psycholinguistic and anthropological theories that jointly predict the probabilities for using overt noun phrase arguments in discourse (“referential density”). Stivers et al. (2009) propose a universal theory of human communication that predicts strong probabilistic patterns in turn-taking, such as response latencies in conversation. But much else is unexplored here: for example, beyond anecdotal observations, we know little about the areal diffusion of discourse and conversational practices although such patterns strike fieldworkers everywhere in the world. Progress in these areas requires

that typology move from abstract grammatical structures to corpus and experimental data as their object of inquiry. The rapid increase in available corpora worldwide bears great promise here.

I would like to conclude this chapter by an overall lesson that I think one can draw from the current state of the art in linguistics: the field is slowly becoming a “normal” science: a science that relies on the same view of causal theories and their testable, statistical effects that has been so extremely successful across many disciplines. This makes it much easier to embed linguistics into interdisciplinary research. However, given the dual importance of functional and event-based theories, interdisciplinary effort cannot and should not be limited to biology and the cognitive sciences but needs to extend to sociology and history as well.

Further Reading

A general appraisal of the state of the art in typology can be found in the journal *Linguistic Typology*, issue 11:1 (2007), and in 2011, the same journal included a special issue (15:2) on current statistical methods. For comparisons of statistical and absolute generalizations see in particular Dryer (1998), Newmeyer (2005), and Bickel (2010a, 2014). The role of explanation in typology is insightfully discussed in Dryer (2006a,b) (and Section 1 is much inspired by this work). The basic idea of a distribution-based, population science approach to typology was introduced in Nichols (1992, 2003), while the focus on diachrony as the key to understanding current distributions goes back to at least the work of Greenberg (1978) (summarized in modern perspective by Croft 2003; also cf. Givón, this volume). Functional theories have a long tradition in typology – see, among many others, for example Hawkins (1994, 2004) for morphosyntax or Blevins (2004) for phonology – while event-based theories have gained substantial ground in typology only since Nichols (1992). Generalized linear models are introduced in most modern statistics textbooks and specifically also in textbooks for linguists (Johnson 2008, Baayen 2008). For applications of the Family Bias Method, see Bickel (2011c, 2013), Bickel et al. (in press-a,-b), and Sinnemäki (2010).

References

- Baayen, R. Harald. 2008. *Analyzing linguistic data: a practical introduction to statistics*. Cambridge: Cambridge University Press.
- Baker, Mark C. 2001. *The atoms of language: the mind's hidden rules of grammar*. Oxford: Oxford University Press.
- Bakker, Dik. 2011. Language sampling. In Jae Jung Song (ed.), *The Oxford Handbook of Language Typology*, 90–127. Oxford: Oxford University Press.
- Biberauer, Theresa, Anders Holmberg & Ian Roberts. 2008. Structure and linearization in disharmonic word orders. *Proceedings of the 26th West Coast Conference on Formal Linguistics* 96–104.
- Biberauer, Theresa, Anders Holmberg & Ian Roberts. 2014. A syntactic universal and its consequences. *Linguistic Inquiry* 45. 169–225.
- Bickel, Balthasar. 2003. Referential density in discourse and syntactic typology. *Language* 79. 708–736.
- Bickel, Balthasar. 2007. Typology in the 21st century: major current developments. *Linguistic Typology* 11. 239–251.
- Bickel, Balthasar. 2010a. Absolute and statistical universals. In Patrick Colm Hogan (ed.), *The Cambridge Encyclopedia of the Language Sciences*, 77–79. Cambridge: Cambridge University Press.
- Bickel, Balthasar. 2010b. Capturing particulars and universals in clause linkage: a multivariate analysis. In Isabelle Bril (ed.), *Clause-hierarchy and clause-linking: the syntax and pragmatics interface*, 51–101. Amsterdam: Benjamins.
- Bickel, Balthasar. 2011a. Grammatical relations typology. In Jae Jung Song (ed.), *The Oxford handbook of language typology*, 399–444. Oxford: Oxford University Press.
- Bickel, Balthasar. 2011b. Multivariate typology and field linguistics: a case study on detransitivization in Kiranti (Sino-Tibetan). In Peter K. Austin, Oliver Bond, David Nathan & Lutz Marten (eds.), *Proc. Conf. Language Documentation and Linguistic Theory* 3, 3–13. London: SOAS.
- Bickel, Balthasar. 2011c. Statistical modeling of language universals. *Linguistic Typology* 15. 401–414.
- Bickel, Balthasar. 2013. Distributional biases in language families. In Balthasar Bickel, Lenore A. Grenoble, David A. Peterson & Alan Timberlake (eds.), *Language typology and historical contingency*, 415–444. Amsterdam: Benjamins.
- Bickel, Balthasar. 2014. Linguistic diversity and universals. In Nicholas J. Enfield, Paul Kockelman & Jack Sidnell (eds.), *The Cambridge Handbook of Linguistic Anthropology*, 102–127. Cambridge: Cambridge University Press.
- Bickel, Balthasar, Kristine Hildebrandt & René Schiering. 2009. The distribution of phonological word domains: a probabilistic typology. In Janet Grijzenhout & Barış Kabak (eds.), *Phonological domains: universals and deviations*, 47–75. Berlin: Mouton de Gruyter.
- Bickel, Balthasar, Giorgio Iemmolo, Taras Zakharko & Alena Witzlack-Makarevich. 2013. Patterns of alignment in verb agreement. In Dik Bakker & Martin Haspelmath (eds.), *Languages across boundaries: studies in memory of Anna Siewierska*, 15–36. Berlin: De Gruyter Mouton.
- Bickel, Balthasar & Johanna Nichols. 2006. Oceania, the Pacific Rim, and the theory of linguistic areas. *Proc. Berkeley Linguistics Society* 32. 3 – 15.
- Bickel, Balthasar & Johanna Nichols. 2009. The geography of case. In Andrej Malchukov & Andrew Spencer (eds.), *The Oxford Handbook of Case*, 479–493. Oxford: Oxford University

- Press.
- Bickel, Balthasar, Alena Witzlack-Makarevich & Taras Zakharko. 2014. Typological evidence against universal effects of referential scales on case alignment. In Ina Bornkessel-Schlesewsky, Andrej Malchukov & Marc Richards (eds.), *Scales: a cross-disciplinary perspective on referential hierarchies*, 7–43. Berlin: De Gruyter Mouton.
- Blevins, Juliette. 2004. *Evolutionary phonology: the emergence of sound patterns*. New York: Cambridge University Press.
- Bresnan, Joan. 2007. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In Sam Featherston & Wolfgang Sternefeld (eds.), *Roots: linguistics in search of its evidential base*, 77–96. Berlin: Mouton de Gruyter.
- Bresnan, Joan & Marilyn Ford. 2010. Predicting syntax: processing dative constructions in American and Australian varieties of English. *Language* 86. 186–213.
- Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, Mass.: MIT Press.
- Chomsky, Noam. 1995. *The minimalist program*. Cambridge, Mass.: MIT Press.
- Chomsky, Noam. 2010. Some simple evo devo theses: how true might they be for language? In Richard K. Larson, Viviane Déprez & Hiroko Yamakido (eds.), *The evolution of human language: biolinguistic perspectives*, 45–62. Cambridge University Press.
- Corbett, Greville G. 2005. The canonical approach in typology. In Zygmunt Frajzyngier, Adam Hodges & David S. Rood (eds.), *Linguistic diversity and language theories*, 25–49. Amsterdam: Benjamins.
- Croft, William. 2003. *Typology and universals*. Cambridge: Cambridge University Press [2nd edition].
- Croft, William & Keith T. Poole. 2008. Inferring universals from grammatical variation: multidimensional scaling for typological analysis. *Theoretical Linguistics* 34. 1–37.
- Cysouw, Michael. 2005. Quantitative methods in typology. In Gabriel Altmann, Reinhard Köhler & R. Piotrowski (eds.), *Quantitative linguistics: an international handbook*, 554–578. Berlin: Mouton de Gruyter.
- Cysouw, Michael. 2010. On the probability distribution of typological frequencies. In Christian Ebert, Gerhard Jäger & Jens Michaelis (eds.), *The Mathematics of Language*, 29–35. Berlin: Springer.
- Cysouw, Michael, Mihai Albu & Andreas Dress. 2008. Analysing feature consistency using similarity matrices. *Language Typology and Universals* 61. 263–279.
- Dediu, Dan. 2011. Are languages really independent from genes? If not, what would a genetic bias affecting language diversity look like? *Human Biology* 83. 279–296.
- Dediu, Dan & Robert D. Ladd. 2007. Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, *ASPM* and *Microcephalin*. *Proceedings of the National Academy of Sciences of the USA* 104. 10944–10949.
- Dixon, R. M. W. 2010–2012. *Basic linguistic theory*. Oxford: Oxford University Press.
- Donohue, Mark, Simon Musgrave, Bronwen Whiting & Søren Wichmann. 2011. Typological feature analysis models linguistic geography. *Language* 87. 369–383.
- Dryer, Matthew S. 1989. Large linguistic areas and language sampling. *Studies in Language* 13. 257–292.
- Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68. 81–138.
- Dryer, Matthew S. 1998. Why statistical universals are better than absolute universals. *Papers from the 33rd Annual Meeting of the Chicago Linguistic Society* 123–145.
- Dryer, Matthew S. 2000. Counting genera vs. counting languages. *Linguistic Typology* 4. 334–350.

- Dryer, Matthew S. 2002. Case distinctions, rich verb agreement, and word order type (comments on Hawkins' paper). *Theoretical Linguistics* 28. 151–157.
- Dryer, Matthew S. 2005a. Order of adposition and noun phrase. In Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie (eds.), *The world atlas of language structures*, 346–349. Oxford University Press.
- Dryer, Matthew S. 2005b. Order of genitive and noun. In Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie (eds.), *The world atlas of language structures*, 350–353. Oxford University Press.
- Dryer, Matthew S. 2005c. Order of subject, object, and verb. In Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie (eds.), *The world atlas of language structures*, 330–341. Oxford University Press.
- Dryer, Matthew S. 2005d. Polar questions. In Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie (eds.), *The world atlas of language structures*, 470–473. Oxford: Oxford University Press.
- Dryer, Matthew S. 2006a. Descriptive theories, explanatory theories, and basic linguistic theory. In Felix Ameka, Alan Dench & Nicholas Evans (eds.), *Catching Language: Issues in Grammar Writing*, Berlin: Mouton de Gruyter.
- Dryer, Matthew S. 2006b. Functionalism and the theory-metalanguage confusion. In Grace Wiebe, Gabby Libben, Tom Priestly, Ron Smyth & Sam Wang (eds.), *Phonology, morphology and the Empirical Imperative: papers in honour of Bruce Derwing*, 27–59. Taipei: Crane.
- Dunn, Michael J., Simon J. Greenhill, Stephen C. Levinson & Russell D. Gray. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473. 79–82.
- Enfield, Nicholas J. 2005. Areal linguistics and Mainland Southeast Asia. *Annual Review of Anthropology* 34. 181–206.
- Evans, Nicholas. 2003. Context, culture, and structuration in the languages of Australia. *Annual Review of Anthropology* 32. 13–40.
- Fitch, W. Tecumseh. 2011. Unity and diversity in human language. *Philosophical Transactions of the Royal Society B* 366. 376–388.
- Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of Language*, 73–113. Cambridge, Mass.: MIT Press.
- Greenberg, Joseph H. 1978. Diachrony, synchrony and language universals. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.), *Universals of human language I: method and theory*, 61–92. Stanford: Stanford University Press.
- Haspelmath, Martin. 1998. How young is Standard Average European? *Language Sciences* 20. 271–287.
- Haspelmath, Martin. 2010. Comparative concepts and descriptive categories: consequences for language description and typology. *Language* 86. 663–687.
- Hauser, Marc D., Noam Chomsky & W. Tecumseh Fitch. 2002. The faculty of language: what it is, who has it, and how did it evolve? *Science* 298. 1569–1579.
- Hawkins, John A. 1983. *Word order universals*. New York: Academic Press.
- Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
- Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

- Heine, Bernd & Tania Kuteva. 2006. *The changing languages of Europe*. Oxford: Oxford University Press.
- Jaeger, T. Florian, Peter Graff, William Croft & Daniel Pontillo. 2011. Mixed effect models for genetic and areal dependencies in linguistic typology. *Linguistic Typology* 15. 281–320.
- Jakobson, Roman. 1931. K xarakteristike evrazijskogo jazykovogo sojuza. In *Selected Writings* 1, 144–201. The Hague: Mouton 1970.
- Johnson, Keith. 2008. *Quantitative methods in linguistics*. London: Blackwell.
- Karlsson, Fred. 2009. Origin and maintenance of clausal embedding complexity. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an evolving variable*, 192–202. Oxford: Oxford University Press.
- Levinson, Stephen C. & Nicholas D. Evans. 2010. Time for a sea-change in linguistics: response to comments on ‘The Myth of Language Universals’. *Lingua* 2733–2758.
- Levinson, Stephen C., Simon J. Greenhill, Russell D. Gray & Michael J. Dunn. 2011. Universal typological dependencies should be detectable in the history of language families. *Linguistic Typology* 15. 509–534.
- Maslova, Elena. 2000. A dynamic approach to the verification of distributional universals. *Linguistic Typology* 4. 307–333.
- Maslova, Elena. 2004. Dinamika tipologičeskix raspredelenij i stabil’nost’ jazykovyx tipov. *Voprosy Jazykoznanija* 2004. 3–16.
- Maslova, Elena & Tatiana Nikitina. 2007. Stochastic universals and dynamics of cross-linguistic distributions: the case of alignment types. Ms. Stanford University, <http://another.summa.net/Publications/Ergativity.pdf>.
- Meyer, David, Achim Zeileis & Kurt Hornik. 2006. The **strucplot** framework: visualizing multi-way contingency tables with **vcd**. *Journal of Statistical Software* 17. 1–48.
- Mithun, Marianne. 2010. The fluidity of recursion and its implications. In Harry G. van der Hulst (ed.), *Recursion and human language*, 17–41. Berlin: De Gruyter Mouton.
- Mous, Maarten. 1997. The position of the object in Tunen. In Rose-Marie Déchaine & Victor Manfredi (eds.), *Object Positions in Benue-Kwa*, 123–137. The Hague: Holland Academic Graphics.
- Newmeyer, Frederick J. 2005. *Possible and probable languages: a generative perspective on linguistic typology*. New York: Oxford University Press.
- Newmeyer, Frederick J. 2010. On comparative concepts and descriptive categories: a reply to Haspelmath. *Language* 86. 688–695.
- Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: The University of Chicago Press.
- Nichols, Johanna. 1998. The Eurasian spread zone and the Indo-European dispersal. In Roger Blench & Matthew Spriggs (eds.), *Archeology and language II: archeological data and linguistic hypotheses*, 220–266. London: Routledge.
- Nichols, Johanna. 2003. Diversity and stability in language. In Richard D. Janda & Brian D. Joseph (eds.), *Handbook of Historical Linguistics*, 283–310. London: Blackwell.
- Nichols, Johanna & Balthasar Bickel. 2009. The AUTOTYP genealogy and geography database: 2009 release. Electronic database, <http://www.uzh.ch/spw/autotyp>.
- Piantadosi, Steven T. & Edward Gibson. 2014. Quantitative standards for absolute linguistic universals. *Cognitive Science* 38. 736–756.
- R Development Core Team. 2014. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing, <http://www.r-project.org>.

- Rigsby, Bruce & Noel Rude. 1996. Sketch of Sahaptin, a Sahaptian language. In Ives Goddard (ed.), *Handbook of North American Indians, Vol. 17: Languages*, 666–692. Washington: Smithsonian Institution.
- Rootsi, S., L.A. Zhivotovsky, M. Baldovic, M. Kayser, I.A. Kutuev, R. Khusainova, M.A. Bermisheva, M. Gubina, S. Fedorova, A.M. Ilumäe, E.K. Khusnutdinova, L.P. Osipova, M. Stoneking, V. Ferak, J. Parik, T. Kivisild, P.A. Underhill & R. Villems. 2007. A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. *European Journal of Human Genetics* 15. 204–211.
- Siewierska, Anna. 1996. Word order type and alignment. *Language Typology and Universals* 49. 149–176.
- Sinnemäki, Kaius. 2010. Word order in zero-marking languages. *Studies in Language* 34. 869–912.
- Stivers, Tanya, Nicholas J. Enfield, Penelope Brown, Christina Englert, Makoto Hayashi, Trine Heinemann, Gertie Hoymann, Federico Rossano, Jan Peter de Ruiter, Kyung-Eun Yoon & Stephen C. Levinson. 2009. Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences* 106. 10587–10592.
- Stobbe, Nina, Gesche Westphal-Fitch, Ulrike Aust & W. Tecumseh Fitch. 2012. Visual artificial grammar learning: comparative research on humans, kea (*Nestor notabilis*) and pigeons (*Columba livia*). *Philosophical Transactions of the Royal Society B: Biological Sciences* 367. 1995–2006.
- Stokhof, Martin & Michiel van Lambalgen. 2011. Abstractions and idealisations: the construction of modern linguistics. *Theoretical Linguistics* 37. 1–26.
- Stoll, Sabine & Balthasar Bickel. 2009. How deep are differences in referential density? In Jiansheng Guo, Elena Lieven, Nancy Budwig, Susan Ervin-Tripp, Keiko Nakamura & Şeyda Özçalışkan (eds.), *Crosslinguistic approaches to the psychology of language: research in the traditions of Dan Slobin*, 543–555. London: Psychology Press.
- Trudgill, Peter. 2011. *Sociolinguistic typology: social determinants of linguistic complexity*. Oxford: Oxford University Press.
- Venables, W. N. & Brian D. Ripley. 2002. *Modern applied statistics with S*. New York: Springer.
- Viti, Carlotta. 2012. *Variation und Wandel in der Syntax der alten indogermanischen Sprachen*. Zurich: UZH Habilitationsschrift.
- Witzlack-Makarevich, Alena. 2011. *Typological variation in grammatical relations*. Leipzig: University of Leipzig PhD dissertation.